Instructions for group assignment: Machine Learning Competition

Overview

In this assignment, you will work collaboratively in groups to develop a predictive model using machine learning techniques. The objective is to estimate the best-performing model for predicting the weight of adults using the dataset provided. This project will allow you to apply your data analytics and machine learning skills to a real-world scenario.

1. Group Formation

- Form groups of 3 to 5 students.
- You are free to choose your group members.
- Groups can be composed of students from either of the two classes in the program.

2. Dataset

You have been provided with an Excel file containing:

- Training Data: This sheet includes the dataset with 617 variables.
- Variable Names: A separate sheet listing the names and descriptions of each variable.

Your task is to use this dataset to develop a machine learning model that predicts WEIGHTLBTC_A (Weight in Pounds).

3. Tools and Methodology

- Use Python to analyse the data and build the predictive model.
- Please limit yourself to the models studied in this course, but you may explore all possible hyperparameter tuning for these models, even if not covered in class.
- Your final model will be evaluated using mean square error.

4. Report Submission

Your group must submit a report explaining how you came up with the final model submitted. Include all algorithms tested, along with any preprocessing, scaling, feature selection and hyperparameter tuning implemented. The length of the report should not exceed 2500 words.

Your group must also submit the python code and the model generated. Instructions on how to do this will follow shortly.

5. Submission Deadline

The final report, along with your Python code, must be submitted before midnight on **9 January 2026**. Late submissions will incur penalties.

Good luck! I look forward to seeing your innovative solutions.